Vision Research 104 (2014) 36–46

Contents lists available at ScienceDirect

Vision Research

journal homepage: www.elsevier.com/locate/visres

Contextual modulation and stimulus selectivity in extrastriate cortex

Matthew R. Krause *, Christopher C. Pack

Department of Neurology and Neurosurgery, Montreal Neurological Institute, McGill University, Montreal, QC, Canada

ARTICLE INFO

Article history:
Received 25 March 2014
Received in revised form 8 October 2014
Available online 3 November 2014

Keywords:
Contextual modulation
Normalization
Surround
Extrastriate cortex
Neurophysiology
Macaque

ABSTRACT

Contextual modulation is observed throughout the visual system, using techniques ranging from single-neuron recordings to behavioral experiments. Its role in generating feature selectivity within the retina and primary visual cortex has been extensively described in the literature. Here, we describe how similar computations can also elaborate feature selectivity in the extrastriate areas of both the dorsal and ventral streams of the primate visual system. We discuss recent work that makes use of normalization models to test specific roles for contextual modulation in visual cortex function. We suggest that contextual modulation renders neuronal populations more selective for naturalistic stimuli. Specifically, we discuss contextual modulation's role in processing optic flow in areas MT and MST and for representing naturally occurring curvature and contours in areas V4 and IT. We also describe how the circuitry that supports contextual modulation is robust to variations in overall input levels. Finally, we describe how this theory relates to other hypothesized roles for contextual modulation.

1. Introduction

Visual information is rarely found in isolation. A typical scene contains many objects, each of which can be defined by its own combination of visual features. Many of these features, such as orientation and spatial frequency, are extracted by dedicated mechanisms in the early visual system. These circuits are thought to generate feature selectivity in part by repeatedly filtering and pooling feedforward inputs. For example, a V1 neuron could develop orientation tuning by selectively pooling the outputs of several circular LGN receptive fields (Hubel & Wiesel, 1962); the LGN receptive fields, in turn, arise from filtering and pooling in the retina and retinal ganglion cells. Since the neurons implementing these operations have small spatial receptive fields and short memories, it may seem like processing should be quite local in space and time.

However, it has long been known that the processing of a visual stimulus is affected by the overall *gestalt*, or context, in which it occurs. The presence of a stimulus, even one that cannot directly drive a neuron's feedforward inputs (e.g., because it is outside the cell's spatial receptive field or its tuning passband), can affect how the cell responds to other stimuli that do engage its feedforward inputs.

A whimsical example of this effect can be found in Quiroga et al. (2005), who recorded the activity of medial temporal lobe (MTL) neurons in human patients while the patients viewed photographs. One neuron, shown in Fig. 1, responded strongly and almost exclusively to photos of the actress Jennifer Aniston (shaded regions in the top row), regardless of the low-level features (e.g., color, edge orientation) that comprise her portrait. This selectivity and invariance is common in high-level cortical areas (Desimone et al., 1984; Tsunoda et al., 2001), but virtually unheard of in lower ones, where cells respond to any stimulus containing an appropriate angle or hue (Hubel & Wiesel, 1968; Leventhal et al., 1995). However, there is virtually no response to images that contain both Aniston and her then-husband, actor Brad Pitt (Fig. 1, top-right), again regardless of the low-level features that make up his appearance. Thus, one might conclude that Brad Pitt's presence suppresses the cells' responses to Jennifer Aniston.

However, other models might also explain these responses. The response pattern might reflect selectivity for a specific, low-level feature (e.g., orientation or color) that happens to be present in all of the Aniston images but none of the Aniston + Pitt images. Or perhaps any stimulus accompanying Aniston, other than the background, leads to suppression. Because the stimulus features that activate MTL neurons are not well understood, there is little basis for estimating the contribution of these different mechanisms.

The difficulty in modeling such complex visual selectivity thus arises from the variety of possible inputs and, in many cases, from a lack of detailed knowledge of the computations performed by cortical neurons selective for complex stimuli. Recent work in this area has attempted to solve this problem by leveraging the

---

* Corresponding author at: Room 888, 3801 rue University, Montreal, QC H3A 2B4, Canada.

E-mail address: matthew.krause@mcgill.ca (M.R. Krause).

http://dx.doi.org/10.1016/j.visres.2014.10.006

available knowledge on low-level visual processing. We suggest that a particularly fruitful approach is to present neurons with a large variety of stimuli that explore, to the extent possible, a wide range of feature combinations. The resulting data are then fit to models that approximate the hierarchical structure of the visual system (Brincat & Connor, 2004; Mineault et al., 2012; Rust et al., 2006). Thus, for example, models of V2 can be framed as performing computations on the outputs of simulated V1 neurons, rather than operating on the raw visual input (Coen-Cagli & Schwartz, 2013). The precise operations that are used typically include feedforward filtering, as well as contextual modulations such as normalization.

2. Contextual modulation in striate cortex

Contextual modulation is typically measured in neurophysiological experiments using a simple paradigm. Investigators isolate a neuron and map its classical receptive field (CRF). They then compare responses to stimuli placed only within the cell's CRF with those that extend beyond its boundaries. When the contents of the CRF are identical in the two conditions, any observed difference is then ascribed to contextual modulation, and the spatial area producing these effects is called the non-classical receptive field (nCRF), or surround.

Although there is some evidence for excitatory contextual modulation (Angelucci & Bressloff, 2006; Bringuier et al., 1999), the net effect of nCRF stimulation is typically suppressive: stimulating the nCRF with large, high contrast stimuli reduces V1 neurons' firing rates by 40–70%, compared to CRF-only stimulation (reviewed in Series et al. (2002)); similar results have also been obtained in extrastriate areas. Several functional roles have been proposed for this modulation, including the following:

- figure-ground segmentation (Allman, Miezin, & McGuinness, 1985),
- redundancy reduction (Atick & Redlich, 1990; Dong & Atick, 1995),
- generation of a sparse code (Vinje & Gallant, 2000),
- firing rate control/metabolic efficiency (Attwell & Laughlin, 2001), and
- noise rejection (Chen, Geisler, & Seidemann, 2006).

These hypotheses all share a common feature: contextual modulation is used to refine existing feature representations that have been generated by other—unspecified, but presumably feedforward—circuitry. Here we review evidence suggesting that contextual modulation can do more, and actually creates neural selectivity for new and complex visual features. There is almost universal agreement that this occurs in the retina, where contextual modulation—implemented through lateral inhibition—converts the absolute luminance information captured by the retina into a new image feature, local contrast (Hartline, 1940; Kuffler, 1953). This review focuses on the consequences of iterating similar mechanisms across multiple visual cortical areas, a topic that has been explored less thoroughly (but see Gautama & Van Hulle, 2001).

Contextual modulation is typically thought to arise from interactions between neurons. These interactions can take several forms. When expressed mathematically as a subtraction of two quantities, akin to the integration of IPSPs and EPSPs, the modulation is usually called opponent inhibition (Hurvich & Jameson, 1957; Reid & Shapley, 1992). Interactions between neighboring bipolar or amacrine cells, for example, are often described using opponent models. When these interactions are expressed using a divisive interaction between neurons—or populations of neurons—the resulting model is usually called a normalization model. These models have a long history in visual neuroscience

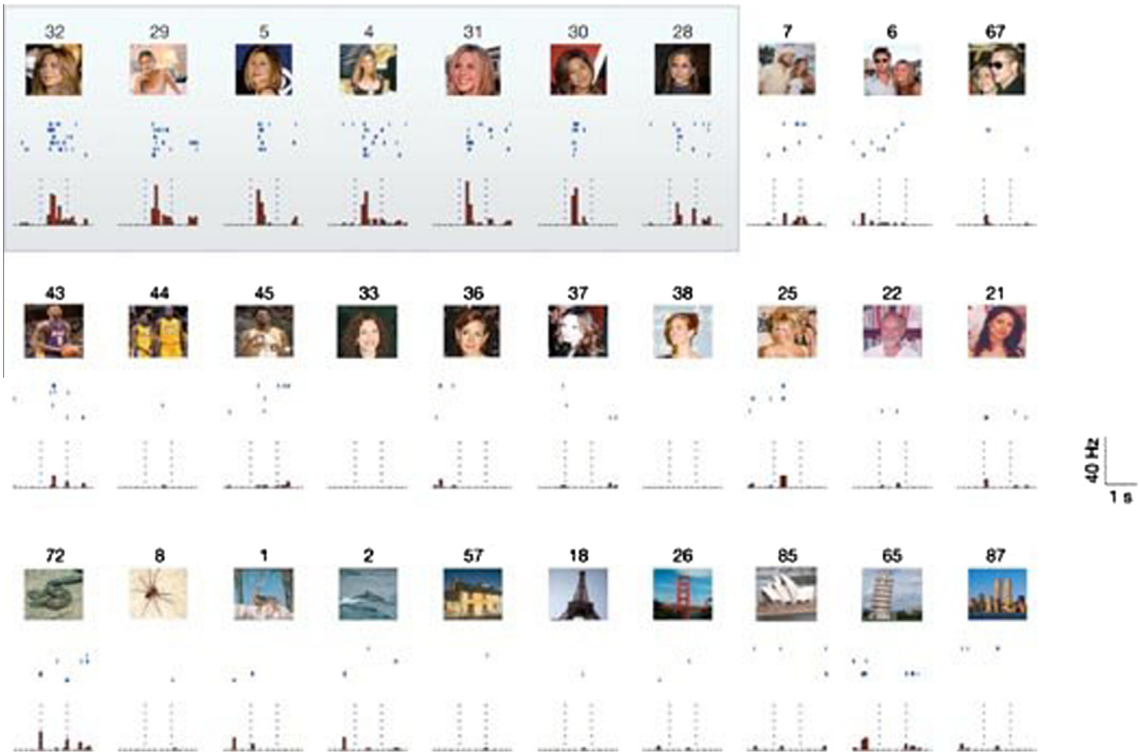

Fig. 1. Do MTL neurons exhibit complex contextual modulation? Quiroga et al. (2005) recorded the activity of a medial temporal lobe (MTL) neuron while human patients viewed images. This neuron responded vigorously (individual trials shown in center; peristimulus histograms shown in the bottom row) whenever the patient saw Jennifer Aniston, but was suppressed whenever Brad Pitt was also in the photograph. However, we know very little about MTL neurons' feature selectivity or the computations they perform that might evoke this suppression. Reprinted by permission from Macmillan Publishers Ltd.: Nature (Quiroga et al., 2005) © 2005.

(Grossberg, 1973; Naka & Rushton, 1966), and some researchers have argued that normalization is a "canonical [form] of neural computation" that can be found across diverse brain areas and species (Carandini & Heeger, 2012). Since they have gained such widespread acceptance in the vision literature, normalization models provide a means of comparing computations across different brain regions.

The basic structure of a normalization model (Grossberg, 1973; Naka & Rushton, 1966) is as follows. The response R_i of a single neuron i is given by:

$$R_i = M\frac{I_i}{\sum_j w_j I_j + \sigma} \tag{1}$$

where M is the maximum firing rate, I_i is the sum of the neuron's excitatory input, σ is a constant, and the I_js represent inputs from other neurons, weighted by the corresponding w_js. The ensemble of I_js and w_js defines the *normalization pool*. Functionally, the normalization pool allows the response of a neuron to be influenced by stimuli, or even non-visual inputs, that by themselves have little or no net influence on the neuron's response. In other words, it provides contextual modulation. As we describe below, the constitution of the normalization pool is crucial for understanding contextual modulation in the visual cortex.

2.1. The normalization model of contrast gain

Early applications of the normalization model to the primate visual cortex (Albrecht & Geisler, 1991; Heeger, 1992) proposed a normalization pool comprised of neurons tuned collectively to all orientations and spatial frequencies. This *untuned* normalization mechanism effectively encodes the total contrast in a local image patch. In this approach, both the feedforward input to the neuron I_i and the total activity in the normalization pool are assumed to be proportional to contrast c. Thus $I_i = cT_i(\theta)$, where T represents the feedforward tuning of the neuron over the space of features θ and k is a free parameter; Eq. (1) then becomes:

$$R_i = M\frac{cT_i(\theta)}{kc + \sigma} \tag{2}$$

As a result of this formulation, there is a nonlinear relationship between contrast and the neural response, but, importantly, the neuron's selectivity for stimulus features, as determined by the ensemble of excitatory input, is independent of contrast. This seems like a desirable property in general, as there is no reason why visual perception of the surrounding environment should change with contrast. Moreover, it is consistent with the frequent observation that the tuning of V1 neurons for stimulus orientation is invariant to changes in contrast (Anderson et al., 2000).

2.2. The normalization model of size tuning

The basic untuned normalization mechanism described above can account for another common property of visual cortex neurons: size tuning. Most V1 neurons in the primate respond more strongly as stimulus size is increased, up to the point at which the stimulus fills the entire classical receptive field. As size is increased further, the response often begins to decrease (Hubel & Wiesel, 1965; Series, Lorenceau, & Fregnac, 2003), typically settling at a level that is less than half that of the response observed at the optimal size. These findings can be accounted for with a simple extension to the normalization model. If the normalization pool includes neurons with receptive fields in a variety of different positions, suppression of the input will be observed for large stimuli, which increase the denominator of Eq. (1), but not the numerator. This phenomenon is often called surround suppression.

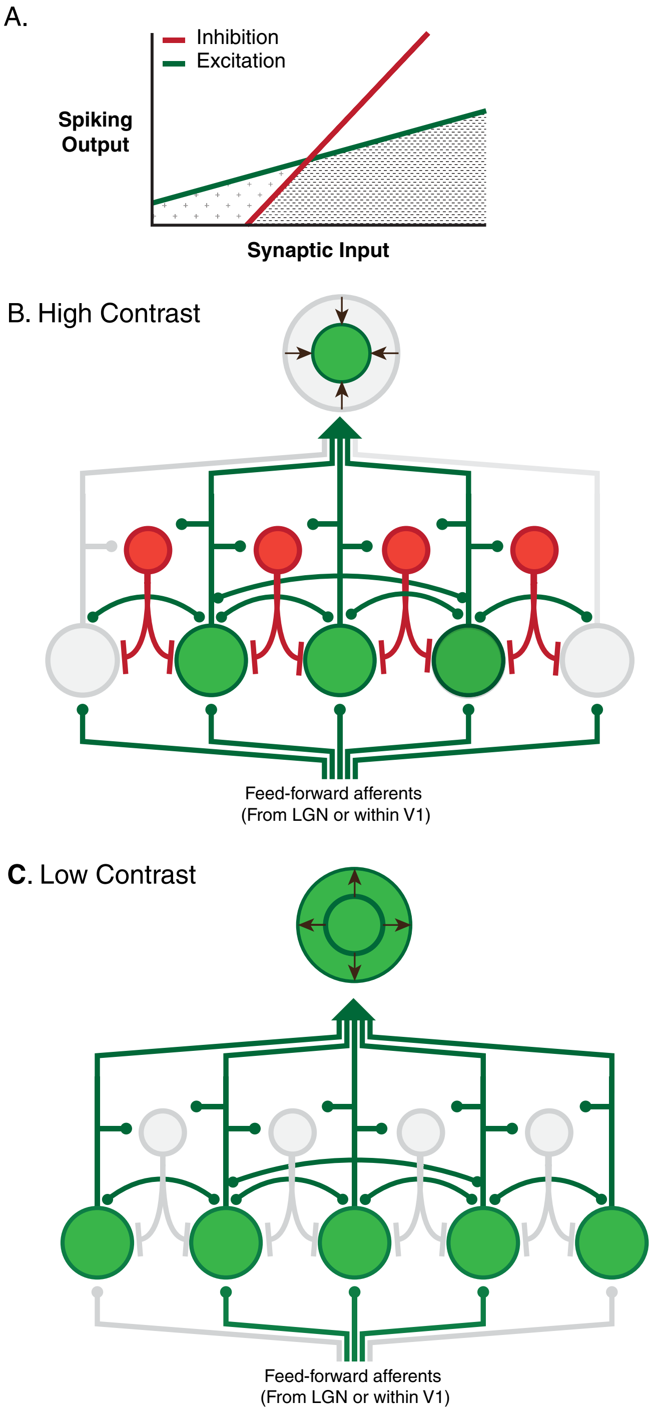

Fig. 2. Circuitry for size tuning/surround suppression in primary visual cortex. (A) Input–output relationships for excitatory pyramidal cells and inhibitory interneurons, redrawn from Nowak et al. (2003). Interneurons have a higher gain than pyramidal cells, but also require more input to begin firing at all. This mismatch produces spatial summation in low contrast conditions where cells receive weak synaptic input, and surround suppression at higher contrasts when they receive stronger synaptic input. (B) Feed-forward input drives both lateral excitatory connections (green), but also activates inhibitory interneurons (red). The interneurons' higher gain allows them to suppress some of the lateral activity, shrinking the cell's receptive field and creating a suppressive surround. (C) Feed-forward input fails to activate the inhibitory interneurons, so excitation spreads via lateral connections. This produces spatial summation and a larger receptive field. The circuits in B and C are simplified versions of the circuit proposed by Angelucci, Levitt, and Lund (2002). Their complete model contains an additional source of inhibition, not shown here, to account for orientation-dependent effects.

Fig. 2 illustrates a circuit that produces size tuning/surround suppression in primary visual cortex.

3. Contextual processing in extrastriate visual cortex

Most theoretical and empirical work on contextual modulation has focused on the primary visual cortex, as outlined above. This makes sense, given the relatively advanced state of knowledge on the anatomy and functional organization of V1 (Angelucci & Bressloff, 2006). However, the role of contextual modulation in vision must also be understood with reference to the extrastriate cortex, a collection of brain regions that are generally thought to be more closely associated with conscious visual perception. In this section, we describe recent progress in understanding the impact of contextual modulation on extrastriate cortical function.

3.1. Motion processing and tuned normalization

Extrastriate cortical regions can, broadly speaking, be assigned to two categories. The first is a dorsal pathway originating in V1 and terminating in the parietal cortex; the second ventral pathway, also originating in V1, projects toward the temporal lobe and hippocampal structures (Felleman & Van Essen, 1991; Mishkin & Ungerleider, 1982).

Of the two pathways, the dorsal (motion) pathway is somewhat better understood, in part because the space of relevant stimuli is more easily parameterized. Since the eyes themselves are almost never still (Otero-Millan et al., 2008), and objects are typically stationary (Stocker & Simoncelli, 2006), most motion encountered by the visual system is due to displacement of the eye in space, either through saccades, smooth pursuit, head rotation, or navigation. The resulting *optic flow* stimuli are thus constrained by the translation and rotation of the eye and by the structure of the surrounding environment. By making simplifying assumptions about the latter, we can reduce the number of motion stimuli to a manageable subspace; neuronal responses can then be studied by modeling the transformation of stimuli within this space into firing rates.

The modeling effort is further simplified by the fact that the visual cortex regions devoted to motion processing in the primate have been studied thoroughly (Born & Bradley, 2005). For example we have a fairly good understanding of the estimation of motion by subpopulations of neurons in the primary visual cortex (V1). The responses of these neurons can be reasonably well approximated with motion energy models (Adelson & Bergen, 1985), which, importantly, can be elaborated upon to include contextual modulation (Tsui et al., 2010). Moreover, the subsequent transformation of V1 outputs into stimulus selectivity in the middle temporal (MT) area is also relatively well-understood (Born & Bradley, 2005). By leveraging the large body of existing literature on these topics, we can examine quantitatively the function of contextual modulation in motion processing.

Fig. 3 illustrates some possible roles for context modulation in motion processing. The small receptive fields of V1 neurons (purple ovals) limit their ability to extract estimates of velocity for larger objects composed of oriented edges. This is known as the "aperture problem" in the vision literature (Marr & Ullman, 1981); it is a specific case of the more general fact that local information is not necessarily indicative of global structure. The aperture problem is particularly useful as a probe of visual processing, because it provides a simple geometric way to dissociate local and global image motion. In the cartoon example in Fig. 3A, the V1 neurons would report purely vertical or horizontal motion (thin arrows), even though the rectangle is moving obliquely (thick arrow). The correct motion direction can thus only be recovered by combining inputs across local motion directions, as is done in MT.

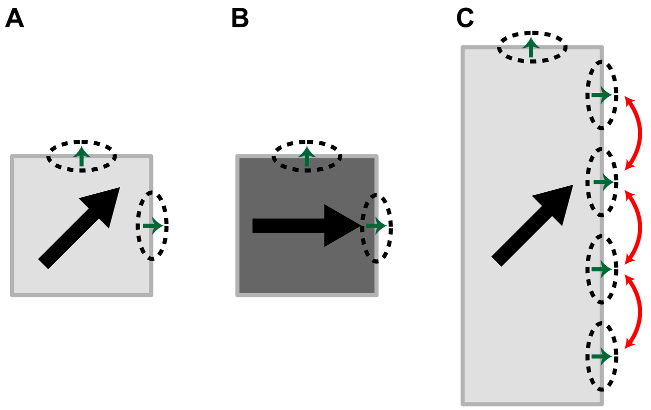

Fig. 3. Contextual modulation is required to recover global motion. V1 neurons can only detect motion in one direction (small green arrows), along the minor axis of their receptive fields. Real motion, however, is less constrained. (A) Averaging the output of two such detectors sometimes correctly recovers an object's motion (large black arrow). (B) The object in B has twice the contrast as the object in A; this activates each V1 subunit twice as much as in A. However, averaging the output of these two detectors would erroneously suggest that object B was moving in the same direction and at the same speed as object A. (C) The asymmetric shape of the object in C activates more rightward neurons than upward neurons, even though the object is moving in the same direction as in A. Mutual suppression between similarly-tuned direction detectors (curved red arrows) ensures that the object is seen moving in the appropriate direction. See Section 3.1 for more details.

Contextual modulation can facilitate the process of combining inputs. If the responses of the V1 neurons were linear in contrast, an MT neuron that summed their outputs would not be able to distinguish between the two stimuli shown in Fig. 3A and B, even though their motion directions differ by 45°. This results from the fact that the contrast of the stimulus in Fig. 3B is twice than that of the contrast in Fig. 3A. Consequently a high-contrast, rightward moving stimulus would elicit the same response as a low-contrast, obliquely moving object. Contrast normalization of the kind implemented by Eq. (2) could ameliorate this problem, as it causes individual V1 neurons' responses to saturate at low contrasts. Consequently the summed response of the pair of V1 neurons to the stimulus in Fig. 3A would be greater than that to the stimulus in Fig. 3B. Thus contrast normalization can have dramatic effects on downstream neuronal selectivity.

Fig. 3C poses a different sort of challenge. The object moves in the same oblique direction as that in Fig. 3A, but its asymmetric shape activates more V1 neurons tuned to rightward motion. Ideally an MT neuron tuned to upward–rightward motion would respond identically to the stimuli in Fig. 3A and C, and this is indeed what generally happens (Tsui et al., 2010). Contextual modulation can contribute to this process: if there is suppression among rightward-tuned V1 neurons, the additional activity due to the longer edge will be countered by additional suppression due to the contextual modulation. Thus the total output of all rightward motion detectors in the vicinity of the object will be largely invariant to stimulus size and contrast (Tsui et al., 2010). This type of stimulus-tuned modulation allows contextual modulation to generate invariances that aid motion processing, as well as other functions (Carandini & Heeger, 2012), some of which are outlined below.

3.1.1. Contextual modulation in the outputs from V1 to MT

The difficulty of determining global motion from local information was originally pointed out on psychophysical grounds (Wallach, 1935, translated by Wuerger, Shapley, & Rubin, 1996).

This and related perceptual phenomena motivated theoretical models (Lorenceau et al., 1993; Rubin & Hochstein, 1993; van den Berg & Noest, 1993) that hypothesized contextual modulation at an early stage, perhaps as early as V1. Common to all these models was the notion that contextual modulation could simplify the integration of motion signals in downstream brain regions. Qualitatively this mechanism would seem to be consistent with the suppressive contextual interactions previously reported in V1 (Dobbins, Zucker, & Cynader, 1987; Hubel & Wiesel, 1965; Pack, Born, & Livingstone, 2003; Versavel, Orban, & Lagae, 1990). Such interactions would complement other mechanisms (e.g., Simoncelli & Heeger, 1998) that are hypothesized to be at work at the level of MT (for detailed reviews, see Born & Bradley, 2005; Bradley & Goyal, 2008).

One simple way to model these contextual interactions is with a normalization model of the form:

$$R_i = M\frac{I_i}{k_1 I_i + \sum_j w_j I_j + \sigma} \tag{3}$$

It is interesting to note that this model is quite similar algebraically to a simple reiteration of the basic normalization model described above (Rust et al., 2006). That is, one obtains the equation above by simply taking the output of Eq. (1) and normalizing it again by itself. The resulting equation has one term in the denominator that depends on all other neurons within the local network (the I_js), as in the untuned contrast normalization model, and another, *tuned* component[1] (I_i). Ringach, Hawken, and Shapley (2003) have also used a similar framework to study orientation tuning in V1. These formulations decouple the excitatory and suppressive components, which allow the model's selectivity to dramatically increase.

Rust et al. (2006) developed a model in which the variability in response properties across the population of MT neurons was assumed to be due to variations in the parameters of a similar equation. They tested the model on data in which MT neurons were stimulated with plaid stimuli comprised of sinusoidal gratings drifting in different directions. The direction of each grating by itself is ambiguous as a consequence of the aperture problem, but the stimulus velocity can be recovered accurately provided that there are two or more gratings present in the stimulus (Adelson & Movshon, 1982). Rust et al. (2006) found that, for neurons that accurately encoded the plaid motion direction (i.e. that exhibited pattern selectivity), the tuned normalization parameter k_1 was particularly important. Models incorporating similar tuned normalization mechanisms also accurately predict MT neurons' responses to moving natural scenes (Nishimoto & Gallant, 2011) and bars (Tsui et al., 2010). Nishimoto and Gallant (2011) augmented a motion-energy model with a static nonlinearity and a divisive normalization stage. These components individually increased the model fit by between 10% and 15%; adding both components led to further (significant) increase in performance and allowed the final model to explain nearly 35% of the explainable variance, on par with contemporary models of other cortical areas.

Tsui et al. (2010) constructed a model in which basic motion energy units (Adelson & Bergen, 1985) received tuned contextual modulation from nearby neurons sharing the same preferences for orientation and motion direction. Importantly, the receptive field locations were offset spatially from those of the excitatory input, instantiating a type of surround suppression often referred to as end-stopping. Beck and Neumann (2011) developed a more detailed model by including separate subpopulations of V1 neurons (complex and end-stopped cells), as well as reciprocal interactions between MT and V1. The tuned normalization in these models is different from many earlier models of contextual modulation, as it implies that the neuron's selectivity for key stimulus features is altered depending on the stimulus context.

One common consequence of tuned normalization is a broadening of stimulus tuning bandwidths. For a strong driving input, the tuned normalization pushes the output of Eq. (1) toward a constant value, effectively flattening the tuning curve around its maximum. This does not happen for weaker inputs, because of the threshold term σ in the denominator. Thus the model predicts that tuning bandwidth should be broader for high-contrast stimuli than for low-contrast stimuli. Tsui et al. (2010) tested this idea explicitly by measuring MT direction tuning for sinusoidal gratings at different stimulus contrasts. The results confirmed the prediction: tuning for grating stimuli was broader at high contrast, when contextual modulation was strongest, than at low contrast, when it was largely absent.

This latter result may seem somewhat puzzling, as the contextual modulation appears to render the neurons less selective for the relevant stimulus feature (motion). However, this decrease in selectivity is stimulus-specific: contextual modulation improves selectivity for stimuli that activate neurons selective for a diversity of local features, as is necessary to recover motion direction (Fig. 3), and as typically occurs during natural vision.

How does contextual modulation improve stimulus selectivity? As mentioned above, in the absence of contextual modulation individual V1 neurons carry very impoverished information about stimulus velocity. Thus any neuron that integrates V1 outputs can recover velocity only by combining information from multiple inputs. The role of contextual modulation is thus to emphasize conjunctions of inputs (Kouh & Poggio, 2008) that are consistent with a single velocity, as proposed in various models of motion integration (e.g., Simoncelli & Heeger, 1998).

3.1.2. Contextual modulation within MT

The function of tuned and untuned contextual modulation has recently been examined in MT. Cui et al. (2013) recorded the responses of MT neurons to various optic flow stimuli and then performed a computational analysis of the roles of different excitatory and suppressive components in generating the observed responses. In particular, they developed models that contained excitatory inputs from V1, along with tuned and untuned suppression that were assumed to arise from lateral connectivity within MT. In general, models that lacked one or both suppressive components failed to account well for the data, suggesting that both contribute to MT responses.

Moreover, the computational work showed that contextual modulation altered the selectivity of MT neurons for optic flow. Without such modulation, the cells generally were capable of representing simple motion trajectories consisting of objects moving in a straight line perpendicular to the line of sight. With contextual modulation, the same neurons became selective for more complex motion patterns that were similar to those encountered in natural vision. These included, for example, patterns consistent with motion in depth or rotation. Such patterns involve motion in different directions in different parts of the visual field; consequently detecting them necessarily involves the detection of conjunctions of inputs sensitive to different local motion directions. Contextual modulation in MT has also been shown to be useful for extracting depth and structure from motion (Buracas & Albright, 1996; Gautama & Van Hulle, 2001).

Thus the results in MT mirror those in V1: contextual modulations alter stimulus selectivity in a manner that allows for more effective processing of naturalistic stimuli. In V1 contextual modulations serve to increase selectivity in MT, and the results of Cui et al. suggest that contextual modulations in

---

[1] As Rust et al. (2006) points out the untuned normalization model shown in Eq. (1) can actually develop a small amount of tuning. However, the tuned model shown in Eq. (3) can exhibit much stronger tuning.

MT might serve to improve selectivity in an area downstream from MT that is selective for optic flow. The medial superior temporal (MST) area would seem to fit the bill, as it integrates the outputs of MT neurons, and is well known for optic flow selectivity.

3.1.3. Contextual modulation in the outputs from MT to MST

Mineault et al. (2012) examined optic flow selectivity in MST, using the same stimulus used by Cui et al. (2013) to probe MT. They also used a similar approach that involved fitting the neuronal responses to a model that transformed the stimulus by various excitatory and suppressive mechanisms. Models with tuned normalization components, similar to those previously proposed for MT (Nishimoto & Gallant, 2011; Rust et al., 2006; Tsui et al., 2010), best predicted MST neurons' responses to optic flow stimuli. This suggests tuned normalization as a repeating motif in cortical processing, though the details of the normalization differed in subtle ways from those found in MT. Specifically, the optimal normalization mechanism was quite local in space, suggesting, somewhat surprisingly, that spatial surround suppression in MT is not necessary to account for selectivity in MST.

The contextual modulation recovered by the Mineault et al. (2012) approach has a functional role that is very similar to that hypothesized for synaptic depression (Abbott et al., 1997) in other systems. In particular, the modulation was spatially very localized, tuned to the same stimulus features as the excitatory inputs, and led to a saturation in the contribution of each individual MT input to a given MST neuron. Previous work in V1 has shown that synaptic depression in thalamocortical synapses can mimic the effects of intracortical normalization, and indeed the mathematical form of the two mechanisms is nearly identical (Carandini, Heeger, & Senn, 2002). An important question for future research then will be to distinguish these two mechanisms experimentally, particularly in the extrastriate cortex.

3.2. Contextual modulation in shape processing

The convergence of functional forms for models of motion processing raises the question of whether similar models might account for visual processing in other domains. Several regions of the extrastriate cortex in primates are associated with shape processing, including occipital areas V2 and V4 and the inferotemporal (IT) cortex. As in the motion pathway, receptive field sizes increase from V1 to V4 to IT, as does the apparent complexity of the stimulus selectivity (Gattass, Gross, & Sandell, 1981; Gattass, Sousa, & Gross, 1988). Measuring the responses to various shapes and using the resulting data to infer the underlying transformations can probe this selectivity.

3.2.1. Contextual modulation in the outputs from V1

The results of various quantitative studies of shape processing support a style of computation similar to that found in the motion processing domain. As with MT neurons, the input to V2 and V4 consists of neurons in earlier stages (such as V1) that are selective for local stimulus features, such as orientation and spatial frequency. As with motion processing, the key elements necessary for shape recognition probably begin with tuned normalization in V1. Specifically, tuned normalization could provide selectivity for local curvature (Dobbins, Zucker, & Cynader, 1987; Hubel & Wiesel, 1965; Versavel, Orban, & Lagae, 1990) a crucial element in shape recognition. Indeed a model in which individual inputs to V4 undergo tuned normalization accounts very well for the selectivity of individual neurons for stimulus shape (Cadieu et al., 2007; Hansen & Neumann, 2008), as proposed originally on psychophysical grounds (Wilson, Wilkinson, & Asaad, 1997).

Similarly, Coen-Cagli and Schwartz (2013) argue that models of V2 should incorporate divisive normalization occurring within the model's inputs from V1. Including this operation allowed the simulated V2 neurons to perform figure-ground tasks better than those that received unnormalized input.

3.2.2. Contextual modulation within V2 and V4

Although existing data show clear evidence for contextual modulation within V2 (Shushruth et al., 2009) and V4 (Pollen et al., 2002), relatively little is known about the precise nature of the modulation. Receptive fields in both areas appear to be assembled from subunits that resemble the receptive fields of earlier areas (V2: Anzai, Peng, & Van Essen, 2007; Nandy et al., 2013; Tao et al., 2012; V4: Pollen et al., 2002). Multiple stimuli placed within V2 and V4 receptive fields often elicited nonlinear response properties. These could be attributed to lateral excitation or inhibition within V4, as was suggested based on theoretical and psychophysical work (Wilson, Krupa, & Wilkinson, 2000). Zanos et al. (2011) found evidence for similar influences in V4 by analyzing functional connectivity among many simultaneously recorded V4 neurons.

Neurons in V2 are sensitive to both orientation and orientation changes/discontinuities occurring within their receptive fields (Ito & Komatsu, 2004; Tao et al., 2012; von der Heydt & Peterhans, 1989). Since this sensitivity develops very rapidly and depends on the cell's own orientation tuning, Schmid, Purpura, and Victor (2014) argue that it must develop within V2. Although a fraction of V1 neurons show similar selectivity, they only do so with delays consistent with feedback from another cortical area (i.e., V2). This functionality may allow V2 neurons to perform figure/ground segmentation and to identify border ownership (Layton, Mingolla, & Yazdanbakhsh, 2012; Zhou, Friedman, & von der Heydt, 2000), perhaps in concert with higher cortical areas.

While V2 neurons have selectivity for orientation combinations or angles (Ito & Komatsu, 2004), neurons in V4 are tuned for local curvature (Gallant et al., 1996; Pasupathy & Connor, 1999). Yau et al. (2013) report that V4 neurons assemble this representation from orientation-tuned inputs, presumably from V1. V4 neurons typically respond for ~100 ms after a stimulus appears. Early portions of the response carry information about the component orientations that comprise local curvature, but curvature selectivity develops gradually over the next 50 ms, consistent with a recurrent origin for V4 curvature selectivity.

3.2.3. Contextual modulation in the outputs from V4 to IT

IT neurons integrate the outputs of neurons in V4, among other areas, to generate quite complex selectivity for object shapes. This selectivity has been probed quantitatively by presenting the neurons with a large battery of static shapes and estimating model components that best account for the observed neuronal responses. This approach is highly analogous to the approaches used in the dorsal stream to study motion.

Brincat and Connor (2004) devised a model in which IT neurons integrate the outputs of V4 neurons tuned for local curvature. The model, illustrated in Fig. 4, was of the form:

$$R_i = \alpha \sum_m w_m I_m + \beta \Pi_m w_m I_m \tag{4}$$

where α and β are weighting factors. The first term in the equation corresponds to a simple linear summation of inputs, whereas the second term involves a multiplication of the same inputs. This second term would be selective to input conjunctions, and part analysis was to determine whether this extra term was justified statistically by the complexity of the data. The results indicated that the multiplicative term improved the fit of the model for many neurons, and in those cells, selectivity for shape was generally higher than in cells that lacked nonlinear responses. Thus the multiplicative interaction among inputs was critical in accounting for shape selectivity in IT neurons.

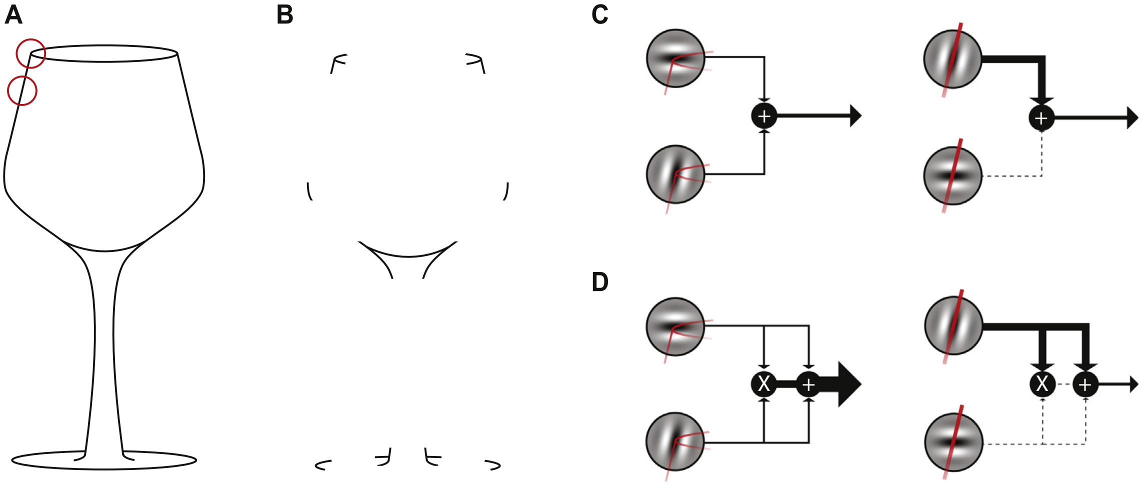

Fig. 4. Curvature and combinations of orientations provide important clues about an objects' shape. (A and B) *Biederman's Cup*. The complete wineglass, shown in Panel A, is also visible in Panel B. Panel B contains less than 10% of the ink in Panel A, but preserves regions containing combinations of orientations and high curvature. This suggests that these features, which are encoded by neurons in V2 and V4/IT, respectively, may be particularly relevant for object recognition. (Drawings after Biederman (1987).) (C) The pooled response of V1-like subunits provides ambiguous information about local structure. On the left, a linear combination of subunits responds to pairs of oriented edges (upper ring, Panel A). However, the circuitry may also respond to a single edge that activates one subunit very strongly (lower ring, Panel A). This ambiguity is analogous to motion processing ambiguity shown in Fig. 3B. (D) Adding a multiplicative interaction, as in Eqs. (4)–(6), ensures that the neuron only responds to stimuli that activate each of its subunits. Stimuli that activate multiple subunits are facilitated, while those that do not fail to activate the cell. See Section 3.2 for more details.

Since the key function of tuned normalization appears to be to render the neurons selective for conjunctions of inputs, it is not surprising that there is a fairly straightforward algebraic relationship between the Brincat and Connor model and the tuned normalization described earlier. Consider the summed output of two neurons that receive identical tuned normalization (Eq. (3)):

$$R = M\left[\frac{I_1}{k_1 I_1 + \sum_j w_j I_j + \sigma} + \frac{I_2}{k_1 I_2 + \sum_j w_j I_j + \sigma}\right] \qquad (5)$$

This equation can be rewritten as:

$$R = M\frac{\alpha(I_1 + I_2) + 2k_1^2 I_1 I_2}{k_1\alpha(I_1 + I_2) + k_1^2 I_1 I_2 + \alpha^2} \qquad (6)$$

where $\alpha = k_2 \sum_j w_j I_j + \sigma$. The numerator is quite similar to the Brincat and Connor model, with separate terms corresponding to summation and multiplication of inputs.[2] The denominator provides tuned normalization, in principle facilitating the detection of conjunctions at the next stage of processing. This suggests that stimulus selectivity in IT can arise from the same types of mechanisms that confer selectivity on neurons in MT and MST. Similar mechanisms have been shown to be at work in the processing of somatosensory motion (Pei et al., 2013) as well.

Contextual modulation may also play a role in developing anterior IT's strong selectivity for specific categories and objects. Winner-take-all operations are an essential component for generating selectivity in neural models of object recognition (Riesenhuber & Poggio, 1999), and Kouh and Poggio (2008) has shown that the same neural circuitry can implement divisive normalization and winner-take-all selection, while also providing a basis for motion-energy or other linear-filtering models. Since so little is known about the parameter space for temporal lobe visual areas, it is difficult to directly fit models to IT data. However, there is some limited evidence for divisive normalization in temporal lobe areas (Zoccolan, Cox, & DiCarlo, 2005).

4. Modulating modulation

The results described in the previous section suggest that tuned normalization is a simple and powerful approach for elaborating feature selectivity, but they also indicate a potential flaw: since the normalization is divisive, even small amounts of noise in the input can produce large distortions. Applying the normalization procedure multiple times, as occurs when signals flow through the visual system, further exacerbates this problem. Since neural noise is often correlated, simply averaging across a large normalization pool cannot resolve this problem (Shadlen & Newsome, 1998). Instead, the visual system uses a different strategy: the threshold for contextual modulation is set higher than for excitatory inputs. As a result, contextual modulation is only active when the corresponding inputs are relatively strong.

4.1. Stimulus properties

Neurons in most visual areas (except perhaps V4: Sani et al., 2013) respond poorly to low contrast stimuli. In these situations, the normalization operation is likely to be dominated by the constant term σ of Eqs. (1)–(3), which would reduce cells' selectivity. Alternately, if the total input to the normalization pool was low, a few noisy neurons could control the normalization pool, thereby amplifying small stochastic fluctuations in the firing rate. Pooling responses—without normalization—avoids these problems and effectively sacrifices some spatial resolution to increase the fidelity of other features. This manifests itself as an increase in receptive field size at low contrast, as occurs throughout the early visual system, including in retinal ganglion cells, the LGN (Nolt, Kumbhani, & Palmer, 2004), primary visual cortex, and area MT (Pack, Hunter, & Born, 2005). Interestingly, this trade-off is probably made anew in each visual area. For example, (Lagae et al., 1989) that the output, but not input, layers of MT show strong surround suppression for random dot patterns, suggesting that this suppression is not inherited from MT's inputs, but instead generated locally or via feedback.

Similar phenomena also affect tuned normalization. In area MT, contextual modulation allows MT neurons to convert the

---

[2] Note that the full expansion of Eq. (5) contains additional terms for all pairwise combinations of inputs. Both Brincat and Connor (2004) and Mineault et al. (2012) found that these terms contributed little to the quality of the model fits.

responses of individual V1 neurons, which are only sensitive to local motion, into an accurate representation of global motion (see Section 3.1 for details). Since this contextual modulation is only engaged at relatively high contrasts, the perceived direction of motion should undergo contrast-dependent changes. This prediction has now been observed many times (Lorenceau et al., 1993; Shiffrar & Lorenceau, 1996; Stone, Watson, & Mulligan, 1990; Weiss, Simoncelli, & Adelson, 2002).

Similar effects may also occur when viewing low luminance stimuli. Rod photoreceptors function mainly in dim light, while brighter stimuli engage the less sensitive cones. The cone subsystem exhibits strong contextual modulation in the retina, but the rod subsystem does not. Rod bipolar cells lack antagonistic center-surround receptive fields (Bloomfield & Xin, 2000), as do light-adapted AII amacrine cells (Xin & Bloomfield, 1999). Similar effects can also be found in cortical areas. Chen et al. (2014) observed that V1 studies of contextual modulation have used a wide range of luminance levels, from a mean luminance of 3 cd/m^2 (Kapadia et al., 1995) up to 75 cd/m^2 (Xing, Yeh, & Shapley, 2009). Those using dimmer stimuli (3–10 cd/m^2) appear to show weaker suppressive effects and more facilitatory effects (see Discussion in Chen et al. (2014)), while brighter stimuli evoked more consistent suppressive effects. This resembles the trade-off made at low contrast: suppressive contextual modulations increase selectivity when ample information is available, but cells trade selectivity for reliability when less information is available.

Similar changes also occur when the visual system is challenged with other sources of uncertainty. Short stimulus presentations lead to inaccurate motion perception (Bowns, 1996; Lorenceau et al., 1993). Huang, Albright, and Stoner (2007) reported that stimulus ambiguity also affects contextual modulation. Solving the aperture problem for contour stimuli requires integration, which engages facilitatory modulation, but when stimulus motion is unambiguous, suppressive contextual modulation helps extract more accurate motion information. They subsequently proposed an MT model, similar to the one described in Fig. 3, which uses changes in input strength to switch between facilitatory and suppressive regimes (Huang, Albright, & Stoner, 2008).

4.2. Attention

Under low contrast or luminance conditions, the contextual information may be so inaccurate that it cannot provide a useful normalization signal, even if the context is otherwise relevant. However, there are also situations where one only wants to consider a small part of the visual input, based on *a priori* information about the locations of behaviorally-relevant objects. Moran and Desimone (1985) demonstrated that spatial attention can selectively gate visual processing by placing a preferred and non-preferred stimulus inside a V4 neuron's receptive field. They cued the animal to attend to one of the two stimuli and found that the neuron responded as if only the attended stimulus was present.

Spatial attention and contextual modulation appear to perform opposite functions: contextual modulation introduces interactions between distant stimuli, while spatial attention isolates the stimuli at one location from the rest. How do these two phenomena interact? Sundberg, Mitchell, and Reynolds (2009) found that spatial attention can partially override contextual modulation. Contextual modulation caused V4 neurons to fire less when stimuli were placed both inside and outside cells' receptive fields. However, attending to a location inside a cell's receptive field decreased contextual suppression by 50%, while attending outside the cell's receptive field increased it by 50%, compared to a baseline condition where attention was very far away.

Burrows and Moore (2009) found that V4 neurons were sensitive to pop-out, a form of contextual modulation that facilitates the detection of stimuli that are dramatically different from the surrounding ones. However, this effect was also abolished when spatial attention was deployed. Contextual modulation apparently affects attended and non-attended stimuli differently. This may reflect a common mechanism for attention and contextual modulation, as proposed by Reynolds and Heeger (2009). Attending to a stimulus increases its apparent contrast (Martinez-Trujillo & Treue, 2002; Reynolds, Pasternak, & Desimone, 2000), or equivalently, increases the input I_i in the normalization operations described above for the attended stimuli. A "flexible normalization pool" model may ensure that the tuned normalization pool is not contaminated by stimuli that are known to be irrelevant for the task at hand (Schwartz & Coen-Cagli, 2013).

5. Conclusion

We have shown that contextual modulation makes neurons more selective for computationally useful image features. In many cases, these normalization models not only fit the data better, but also create selectivity for qualitatively different features than the models' feedforward inputs, as in the MT and MST data described above.

This view of surround suppression is somewhat different from its treatment in the psychophysical literature, which treats the surround as an untuned source of suppression. For example, Tadin et al. (2003) measured motion-detection thresholds in human observers viewing small and large stimuli. These thresholds increased with the size of the stimuli: subjects performed worse at larger sizes, even though the larger stimuli contain more information. They argued that this change matches the activity of individual MT neurons, which fire less in response to large stimuli. Similarly, Zenger-Landolt and Heeger (2003) had subjects perform a contrast-matching paradigm using large and small stimuli, and found that subjects performed more poorly at larger sizes, which parallels the reduced BOLD activity they observed in visual areas when subjects viewed large stimuli.

Changes in single-cell or population activity, however, may not necessarily produce behavioral impairments. Vinje and Gallant (2002) performed an information-theoretic analysis showing that V1 neurons transmit more information about the stimulus when contextually modulated. Crucially, the transmission rate (expressed as bits/second) increases, even as the firing rate (spikes/second) decreases, suggesting that contextual modulation increases the amount of information sent to downstream areas. Thus, contextual modulation may not necessarily impair all visually guided behavior, even if it alters neural and perceptual selectivity in a way that hinders behavioral performance on very specific and synthetic laboratory tasks. Gepshtein, Lesmes, and Albright (2013) recently demonstrated that adaptation, while detrimental under specific experimental conditions, actually represents an optimal allocation of sensory resources in the general case. Understanding contextual modulation's behavioral effects under naturalistic conditions may similarly prove to be a fruitful avenue for future research.

Contextual modulation has also been argued to create sparse or efficient codes for individual neurons or the entire population (Vinje & Gallant, 2000; Willmore, Mazer, & Gallant, 2011). Such an approach seems entirely consistent with the idea of elaborating feature selectivity (Carlson et al., 2011). One approach to developing efficient codes starts with Attneave (1954)'s observation that visual input contains multiple levels of redundancy. Low-level image features such as color change slowly within an object, and object boundaries themselves can be accurately approximated from only a small amount of the input. Accordingly, the visual system should generate representations that minimize the total amount of redundancy in the input (Barlow, 1972; van Hateren, 1992). Early visual areas may whiten the input by removing local spatiotemporal correlations (Atick & Redlich, 1990; Dong & Atick, 1995). Repeated applications of this decorrelation procedure on successively higher-order statistics of the input can produce new feature representations (Schwartz & Simoncelli, 2001). For example, decorrelation of an orientation map might yield local curvature estimates. Karklin and Lewicki have proposed similar hierarchical models (Karklin & Lewicki, 2005; Karklin & Lewicki, 2009) of early visual areas. These models can leverage statistical regularity in their inputs to create selectivity for increasing abstract stimuli and have successfully reproduced complex cell-like phenomena.

5.1. Biophysics of contextual modulation

Since these models contain excitatory and suppressive components, it is tempting to assume that they must map directly onto excitatory and inhibitory neurons in the brain. However, this is not necessarily true. Divisive normalization can be accomplished by a variety of mechanisms (Silver, 2010), including purely feedforward circuitry (Cybenko, 1989; Heeger, 1992), shunting inhibition (Reichardt, Poggio, & Hausen, 1983), or nonlinearities in the cells' inputs (Abbott et al., 1997). Nurminen and Angelucci (2014) discuss the biophysical mechanisms for contextual modulation elsewhere in this special issue.

5.2. Life without contextual modulation

Contextual modulation is incredibly prevalent in the brain. In addition to the visual areas reviewed here, there is evidence for contextual effects in brain areas representing purely abstract concepts such as location (Franconeri, Alvarez, & Cavanagh, 2013) and value (Louie & Glimcher, 2012).

We searched the literature for examples where contextual modulation was notably absent and found very little. However, Thoen et al. (2014) recently reported that the mantis shrimp does not use contextual interactions to generate its representation of color. Every other animal that has been studied—including butterflies, goldfish, and primates—appears to rely on suppressive interactions between photoreceptors to extract color information. However, the mantis shrimp appear to have adopted a motor strategy that uses scanning eye movements to drag each type of photoreceptor across the same portion of the scene (Land et al., 1990). This creates a temporal context, replacing the circuitry-based contextual modulation found in other animals, suggesting that if contextual modulation did not exist, it would still be necessary to invent it.

Acknowledgments

This work was funded by grants to C.C.P. from the MDEIE and from the Canadian Institutes of Health Research (MOP-115178).

References

Abbott, L. F., Varela, J. A., Sen, K., & Nelson, S. B. (1997). Synaptic depression and cortical gain control. *Science, 275*(5297), 220–224.

Adelson, E. H., & Bergen, J. R. (1985). Spatiotemporal energy models for the perception of motion. *Journal of the Optical Society of America A – Optics & Image Science, 2*(2), 284–299.

Adelson, E. H., & Movshon, J. A. (1982). Phenomenal coherence of moving visual patterns. *Nature, 300*(5892), 523–525.

Albrecht, D. G., & Geisler, W. S. (1991). Motion selectivity and the contrast-response function of simple cells in the visual cortex. *Visual Neuroscience, 7*(6), 531–546.

Allman, J., Miezin, F., & McGuinness, E. (1985). Direction- and velocity-specific responses from beyond the classical receptive field in the middle temporal visual area (MT). *Perception, 14*(2), 105–126.

Anderson, J. S., Lampl, I., Gillespie, D. C., & Ferster, D. (2000). The contribution of noise to contrast invariance of orientation tuning in cat visual cortex. *Science, 290*(5498), 1968–1972.

Angelucci, A., & Bressloff, P. C. (2006). Contribution of feedforward, lateral and feedback connections to the classical receptive field center and extra-classical receptive field surround of primate V1 neurons. *Progress in Brain Research, 154*, 93–120. http://dx.doi.org/10.1016/S0079-6123(06)54005-1.

Angelucci, A., Levitt, J. B., & Lund, J. S. (2002). Anatomical origins of the classical receptive field and modulatory surround field of single neurons in macaque visual cortical area V1. *Progress in Brain Research, 136*, 373–388.

Anzai, A., Peng, X., & Van Essen, D. C. (2007). Neurons in monkey visual area V2 encode combinations of orientations. *Nature Neuroscience, 10*(10), 1313–1321. http://dx.doi.org/10.1038/nn1975.

Atick, J. J., & Redlich, A. N. (1990). Towards a theory of early visual processing. *Neural Computation, 2*(3), 308–320. http://dx.doi.org/10.1162/neco.1990.2.3.308.

Attneave, F. (1954). Some informational aspects of visual perception. *Psychological Review, 61*(3), 183–193.

Attwell, D., & Laughlin, S. B. (2001). An energy budget for signaling in the grey matter of the brain. *Journal of Cerebral Blood Flow and Metabolism, 21*(10), 1133–1145. http://dx.doi.org/10.1097/00004647-200110000-00001.

Barlow, H. B. (1972). Single units and sensation: A neuron doctrine for perceptual psychology? *Perception, 1*(4), 371–394.

Beck, C., & Neumann, H. (2011). Combining feature selection and integration – A neural model for MT motion selectivity. *PLoS ONE, 6*(7), e21254. http://dx.doi.org/10.1371/journal.pone.0021254.

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review, 94*(2), 115–147.

Bloomfield, S. A., & Xin, D. (2000). Surround inhibition of mammalian AII amacrine cells is generated in the proximal retina. *Journal of Physiology, 523*(Pt 3), 771–783.

Born, R. T., & Bradley, D. C. (2005). Structure and function of visual area MT. *Annual Review of Neuroscience, 28*, 157–189. http://dx.doi.org/10.1146/annurev.neuro.26.041002.131052.

Bowns, L. (1996). Evidence for a feature tracking explanation of why type II plaids move in the vector sum direction at short durations. *Vision Research, 36*(22), 3685–3694.

Bradley, D. C., & Goyal, M. S. (2008). Velocity computation in the primate visual system. *Nature Reviews Neuroscience, 9*(9), 686–695. http://dx.doi.org/10.1038/nrn2472.

Brincat, S. L., & Connor, C. E. (2004). Underlying principles of visual shape selectivity in posterior inferotemporal cortex. *Nature Neuroscience, 7*(8), 880–886. http://dx.doi.org/10.1038/nn1278.

Bringuier, V., Chavane, F., Glaeser, L., & Fregnac, Y. (1999). Horizontal propagation of visual activity in the synaptic integration field of area 17 neurons. *Science, 283*(5402), 695–699.

Buracas, G. T., & Albright, T. D. (1996). Contribution of area MT to perception of three-dimensional shape: A computational study. *Vision Research, 36*(6), 869–887.

Burrows, B. E., & Moore, T. (2009). Influence and limitations of popout in the selection of salient visual stimuli by area V4 neurons. *Journal of Neuroscience, 29*(48), 15169–15177. http://dx.doi.org/10.1523/JNEUROSCI.3710-09.2009.

Cadieu, C., Kouh, M., Pasupathy, A., Connor, C. E., Riesenhuber, M., & Poggio, T. (2007). A model of V4 shape selectivity and invariance. *Journal of Neurophysiology, 98*(3), 1733–1750. http://dx.doi.org/10.1152/jn.01265.2006.

Carandini, M., & Heeger, D. J. (2012). Normalization as a canonical neural computation. *Nature Reviews Neuroscience, 13*(1), 51–62. http://dx.doi.org/10.1038/nrn3136.

Carandini, M., Heeger, D. J., & Senn, W. (2002). A synaptic explanation of suppression in visual cortex. *Journal of Neuroscience, 22*(22), 10053–10065.

Carlson, E. T., Rasquinha, R. J., Zhang, K., & Connor, C. E. (2011). A sparse object coding scheme in area V4. *Current Biology, 21*(4), 288–293. http://dx.doi.org/10.1016/j.cub.2011.01.013.

Chen, Y., Geisler, W. S., & Seidemann, E. (2006). Optimal decoding of correlated neural population responses in the primate visual cortex. *Nature Neuroscience, 9*(11), 1412–1420. http://dx.doi.org/10.1038/nn1792.

Chen, K., Song, X. M., Dai, Z. Q., Yin, J. J., Xu, X. Z., & Li, C. Y. (2014). The spatial summation characteristics of three categories of V1 neurons differing in non-classical receptive field modulation properties. *Vision Research, 96C*, 87–95. http://dx.doi.org/10.1016/j.visres.2014.01.011.

Coen-Cagli, R., & Schwartz, O. (2013). The impact on midlevel vision of statistically optimal divisive normalization in V1. *Journal of Vision, 13*(8). http://dx.doi.org/10.1167/13.8.13.

Cui, Y., Liu, L. D., Khawaja, F. A., Pack, C. C., & Butts, D. A. (2013). Diverse suppressive influences in area MT and selectivity to complex motion features. *Journal of Neuroscience, 33*(42), 16715–16728. http://dx.doi.org/10.1523/JNEUROSCI.0203-13.2013.

Cybenko, G. (1989). Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals and Systems, 2*(4), 303–314. http://dx.doi.org/10.1007/BF02551274.

Desimone, R., Albright, T. D., Gross, C. G., & Bruce, C. (1984). Stimulus-selective properties of inferior temporal neurons in the macaque. *Journal of Neuroscience, 4*(8), 2051–2062.

Dobbins, A., Zucker, S. W., & Cynader, M. S. (1987). Endstopped neurons in the visual cortex as a substrate for calculating curvature. *Nature, 329*(6138), 438–441. http://dx.doi.org/10.1038/329438a0.

Dong, D. W., & Atick, J. J. (1995). Temporal decorrelation – A theory of lagged and nonlagged responses in the lateral geniculate-nucleus. *Network-Computation in Neural Systems, 6*(2), 159–178. http://dx.doi.org/10.1088/0954-898x/6/2/003.

Felleman, D. J., & Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex, 1*(1), 1–47.

Franconeri, S. L., Alvarez, G. A., & Cavanagh, P. (2013). Flexible cognitive resources: Competitive content maps for attention and memory. *Trends in Cognitive Sciences, 17*(3), 134–141. http://dx.doi.org/10.1016/j.tics.2013.01.010.

Gallant, J. L., Connor, C. E., Rakshit, S., Lewis, J. W., & Van Essen, D. C. (1996). Neural responses to polar, hyperbolic, and Cartesian gratings in area V4 of the macaque monkey. *Journal of Neurophysiology, 76*(4), 2718–2739.

Gattass, R., Gross, C. G., & Sandell, J. H. (1981). Visual topography of V2 in the macaque. *Journal of Comparative Neurology, 201*(4), 519–539. http://dx.doi.org/10.1002/cne.902010405.

Gattass, R., Sousa, A. P., & Gross, C. G. (1988). Visuotopic organization and extent of V3 and V4 of the macaque. *Journal of Neuroscience, 8*(6), 1831–1845.

Gautama, T., & Van Hulle, M. M. (2001). Function of center-surround antagonism for motion in visual area MT/V5: A modeling study. *Vision Research, 41*(28), 3917–3930.

Gepshtein, S., Lesmes, L. A., & Albright, T. D. (2013). Sensory adaptation as optimal resource allocation. *Proceedings of the National Academy of Sciences of the United States of America, 110*(11), 4368–4373. http://dx.doi.org/10.1073/pnas.1204109110.

Grossberg, S. (1973). Contour enhancement, short-term memory, and constancies in reverberating neural networks. *Studies in Applied Mathematics, 52*(3), 213–257.

Hansen, T., & Neumann, H. (2008). A recurrent model of contour integration in primary visual cortex. *Journal of Vision, 8*(8), 1–25. http://dx.doi.org/10.1167/8.8.8.

Hartline, H. K. (1940). The receptive fields of optic nerve fibers. *American Journal of Physiology, 130*(4), 0690–0699.

Heeger, D. J. (1992). Normalization of cell responses in cat striate cortex. *Visual Neuroscience, 9*(2), 181–197.

Huang, X., Albright, T. D., & Stoner, G. R. (2007). Adaptive surround modulation in cortical area MT. *Neuron, 53*(5), 761–770. http://dx.doi.org/10.1016/j.neuron.2007.01.032.

Huang, X., Albright, T. D., & Stoner, G. R. (2008). Stimulus dependency and mechanisms of surround modulation in cortical area MT. *Journal of Neuroscience, 28*(51), 13889–13906. http://dx.doi.org/10.1523/JNEUROSCI.1946-08.2008.

Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *Journal of Physiology, 160*, 106–154.

Hubel, D. H., & Wiesel, T. N. (1965). Receptive fields and functional architecture in two nonstriate visual areas (18 and 19) of the cat. *Journal of Neurophysiology, 28*, 229–289.

Hubel, D. H., & Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *Journal of Physiology, 195*(1), 215–243.

Hurvich, L. M., & Jameson, D. (1957). An opponent-process theory of color vision. *Psychological Review, 64*(Part 1(6)), 384–404.

Ito, M., & Komatsu, H. (2004). Representation of angles embedded within contour stimuli in area V2 of macaque monkeys. *Journal of Neuroscience, 24*(13), 3313–3324. http://dx.doi.org/10.1523/JNEUROSCI.4364-03.2004.

Kapadia, M. K., Ito, M., Gilbert, C. D., & Westheimer, G. (1995). Improvement in visual sensitivity by changes in local context: Parallel studies in human observers and in V1 of alert monkeys. *Neuron, 15*(4), 843–856.

Karklin, Y., & Lewicki, M. S. (2005). A hierarchical Bayesian model for learning nonlinear statistical regularities in nonstationary natural signals. *Neural Computation, 17*(2), 397–423. http://dx.doi.org/10.1162/0899766053011474.

Karklin, Y., & Lewicki, M. S. (2009). Emergence of complex cell properties by learning to generalize in natural scenes. *Nature, 457*(7225), 83–86. http://dx.doi.org/10.1038/nature07481.

Kouh, M., & Poggio, T. (2008). A canonical neural circuit for cortical nonlinear operations. *Neural Computation, 20*(6), 1427–1451. http://dx.doi.org/10.1162/neco.2008.02-07-466.

Kuffler, S. W. (1953). Discharge patterns and functional organization of mammalian retina. *Journal of Neurophysiology, 16*(1), 37–68.

Lagae, L., Gulyas, B., Raiguel, S., & Orban, G. A. (1989). Laminar analysis of motion information processing in macaque V5. *Brain Research, 496*(1–2), 361–367.

Land, M. F., Marshall, J. N., Brownless, D., & Cronin, T. W. (1990). The eye-movements of the mantis shrimp *Odontodactylus-scyllarus* (Crustacea, Stomatopoda). *Journal of Comparative Physiology A – Sensory Neural and Behavioral Physiology, 167*(2), 155–166.

Layton, O. W., Mingolla, E., & Yazdanbakhsh, A. (2012). Dynamic coding of border-ownership in visual cortex. *Journal of Vision, 12*(13), 8. http://dx.doi.org/10.1167/12.13.8.

Leventhal, A. G., Thompson, K. G., Liu, D., Zhou, Y., & Ault, S. J. (1995). Concomitant sensitivity to orientation, direction, and color of cells in layers 2, 3, and 4 of monkey striate cortex. *Journal of Neuroscience, 15*(3 Pt 1), 1808–1818.

Lorenceau, J., Shiffrar, M., Wells, N., & Castet, E. (1993). Different motion sensitive units are involved in recovering the direction of moving lines. *Vision Research, 33*(9), 1207–1217.

Louie, K., & Glimcher, P. W. (2012). Efficient coding and the neural representation of value. *Annals of the New York Academy of Sciences, 1251*, 13–32. http://dx.doi.org/10.1111/j.1749-6632.2012.06496.x.

Marr, D., & Ullman, S. (1981). Directional selectivity and its use in early visual processing. *Proceedings of the Royal Society of London. Series B: Biological Sciences, 211*(1183), 151–180.

Martinez-Trujillo, J., & Treue, S. (2002). Attentional modulation strength in cortical area MT depends on stimulus contrast. *Neuron, 35*(2), 365–370.

Mineault, P. J., Khawaja, F. A., Butts, D. A., & Pack, C. C. (2012). Hierarchical processing of complex motion along the primate dorsal visual pathway. *Proceedings of the National Academy of Sciences of the United States of America, 109*(16), E972–E980. http://dx.doi.org/10.1073/pnas.1115685109.

Mishkin, M., & Ungerleider, L. G. (1982). Contribution of striate inputs to the visuospatial functions of parieto-preoccipital cortex in monkeys. *Behavioural Brain Research, 6*(1), 57–77.

Moran, J., & Desimone, R. (1985). Selective attention gates visual processing in the extrastriate cortex. *Science, 229*(4715), 782–784.

Naka, K. I., & Rushton, W. A. (1966). An attempt to analyse colour reception by electrophysiology. *Journal of Physiology, 185*(3), 556–586.

Nandy, A. S., Sharpee, T. O., Reynolds, J. H., & Mitchell, J. F. (2013). The fine structure of shape tuning in area V4. *Neuron, 78*(6), 1102–1115. http://dx.doi.org/10.1016/j.neuron.2013.04.016.

Nishimoto, S., & Gallant, J. L. (2011). A three-dimensional spatiotemporal receptive field model explains responses of area MT neurons to naturalistic movies. *Journal of Neuroscience, 31*(41), 14551–14564. http://dx.doi.org/10.1523/JNEUROSCI.6801-10.2011.

Nolt, M. J., Kumbhani, R. D., & Palmer, L. A. (2004). Contrast-dependent spatial summation in the lateral geniculate nucleus and retina of the cat. *Journal of Neurophysiology, 92*(3), 1708–1717. http://dx.doi.org/10.1152/jn.00176.2004.

Nowak, L. G., Azouz, R., Sanchez-Vives, M. V., Gray, C. M., & McCormick, D. A. (2003). Electrophysiological classes of cat primary visual cortical neurons in vivo as revealed by quantitative analyses. *Journal of Neurophysiology, 89*(3), 1541–1566. http://dx.doi.org/10.1152/jn.00580.2002.

Nurminen, L., & Angelucci, A. (2014). Multiple components of surround modulation in primary visual cortex: multiple neural circuits with multiple functions? *Vision Research, 104*, 47–56.

Otero-Millan, J., Troncoso, X. G., Macknik, S. L., Serrano-Pedraza, I., & Martinez-Conde, S. (2008). Saccades and microsaccades during visual fixation, exploration, and search: Foundations for a common saccadic generator. *Journal of Vision, 8*(14). http://dx.doi.org/10.1167/8.14.21 (Article 21).

Pack, C. C., Born, R. T., & Livingstone, M. S. (2003). Two-dimensional substructure of stereo and motion interactions in macaque visual cortex. *Neuron, 37*(3), 525–535.

Pack, C. C., Hunter, J. N., & Born, R. T. (2005). Contrast dependence of suppressive influences in cortical area MT of alert macaque. *Journal of Neurophysiology, 93*(3), 1809–1815. http://dx.doi.org/10.1152/jn.00629.2004.

Pasupathy, A., & Connor, C. E. (1999). Responses to contour features in macaque area V4. *Journal of Neurophysiology, 82*(5), 2490–2502.

Pei, Y. C., Chang, T. Y., Lee, T. C., Saha, S., Lai, H. Y., Gomez-Ramirez, M., et al. (2013). Cross-modal sensory integration of visual-tactile motion information: Instrument design and human psychophysics. *Sensors (Basel), 13*(6), 7212–7223. http://dx.doi.org/10.3390/s130607212.

Pollen, D. A., Przybyszewski, A. W., Rubin, M. A., & Foote, W. (2002). Spatial receptive field organization of macaque V4 neurons. *Cerebral Cortex, 12*(6), 601–616.

Quiroga, R. Q., Reddy, L., Kreiman, G., Koch, C., & Fried, I. (2005). Invariant visual representation by single neurons in the human brain. *Nature, 435*(7045), 1102–1107. http://dx.doi.org/10.1038/nature03687.

Reichardt, W., Poggio, T., & Hausen, K. (1983). Figure-ground discrimination by relative movement in the visual-system of the fly. 2. Towards the neural circuitry. *Biological Cybernetics, 46*, 1–30. http://dx.doi.org/10.1007/Bf00595226.

Reid, R. C., & Shapley, R. M. (1992). Spatial structure of cone inputs to receptive fields in primate lateral geniculate nucleus. *Nature, 356*(6371), 716–718. http://dx.doi.org/10.1038/356716a0.

Reynolds, J. H., & Heeger, D. J. (2009). The normalization model of attention. *Neuron, 61*(2), 168–185. http://dx.doi.org/10.1016/j.neuron.2009.01.002.

Reynolds, J. H., Pasternak, T., & Desimone, R. (2000). Attention increases sensitivity of V4 neurons. *Neuron, 26*(3), 703–714.

Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience, 2*(11), 1019–1025. http://dx.doi.org/10.1038/14819.

Ringach, D. L., Hawken, M. J., & Shapley, R. (2003). Dynamics of orientation tuning in macaque V1: The role of global and tuned suppression. *Journal of Neurophysiology, 90*(1), 342–352. http://dx.doi.org/10.1152/jn.01018.2002.

Rubin, N., & Hochstein, S. (1993). Isolating the effect of one-dimensional motion signals on the perceived direction of moving two-dimensional objects. *Vision Research, 33*(10), 1385–1396.

Rust, N. C., Mante, V., Simoncelli, E. P., & Movshon, J. A. (2006). How MT cells analyze the motion of visual patterns. *Nature Neuroscience, 9*(11), 1421–1431. http://dx.doi.org/10.1038/nn1786.

Sani, I., Santandrea, E., Golzar, A., Morrone, M. C., & Chelazzi, L. (2013). Selective tuning for contrast in macaque area V4. *Journal of Neuroscience, 33*(47), 18583–18596. http://dx.doi.org/10.1523/JNEUROSCI.3465-13.2013.

Schmid, A. M., Purpura, K. P., & Victor, J. D. (2014). Responses to orientation discontinuities in V1 and V2: Physiological dissociations and functional implications. *Journal of Neuroscience, 34*(10), 3559–3578. http://dx.doi.org/10.1523/JNEUROSCI.2293-13.2014.

Schwartz, O., & Coen-Cagli, R. (2013). Visual attention and flexible normalization pools. *Journal of Vision, 13*(1). http://dx.doi.org/10.1167/13.1.25.

Schwartz, O., & Simoncelli, E. P. (2001). Natural signal statistics and sensory gain control. *Nature Neuroscience, 4*(8), 819–825. http://dx.doi.org/10.1038/90526.

Series, P., Georges, S., Lorenceau, J., & Fregnac, Y. (2002). Orientation dependent modulation of apparent speed: A model based on the dynamics of feed-forward and horizontal connectivity in V1 cortex. *Vision Research, 42*(25), 2781–2797.

Series, P., Lorenceau, J., & Fregnac, Y. (2003). The "silent" surround of V1 receptive fields: Theory and experiments. *Journal of Physiology – Paris, 97*(4–6), 453–474. http://dx.doi.org/10.1016/j.jphysparis.2004.01.023.

Shadlen, M. N., & Newsome, W. T. (1998). The variable discharge of cortical neurons: Implications for connectivity, computation, and information coding. *Journal of Neuroscience, 18*(10), 3870–3896.

Shiffrar, M., & Lorenceau, J. (1996). Increased motion linking across edges with decreased luminance contrast, edge width and duration. *Vision Research, 36*(14), 2061–2067.

Shushruth, S., Ichida, J. M., Levitt, J. B., & Angelucci, A. (2009). Comparison of spatial summation properties of neurons in macaque V1 and V2. *Journal of Neurophysiology, 102*(4), 2069–2083. http://dx.doi.org/10.1152/jn.00512.2009.

Silver, R. A. (2010). Neuronal arithmetic. *Nature Reviews Neuroscience, 11*(7), 474–489. http://dx.doi.org/10.1038/nrn2864.

Simoncelli, E. P., & Heeger, D. J. (1998). A model of neuronal responses in visual area MT. *Vision Research, 38*(5), 743–761.

Stocker, A. A., & Simoncelli, E. P. (2006). Noise characteristics and prior expectations in human visual speed perception. *Nature Neuroscience, 9*(4), 578–585. http://dx.doi.org/10.1038/nn1669.

Stone, L. S., Watson, A. B., & Mulligan, J. B. (1990). Effect of contrast on the perceived direction of a moving plaid. *Vision Research, 30*(7), 1049–1067.

Sundberg, K. A., Mitchell, J. F., & Reynolds, J. H. (2009). Spatial attention modulates center-surround interactions in macaque visual area V4. *Neuron, 61*(6), 952–963. http://dx.doi.org/10.1016/j.neuron.2009.02.023.

Tadin, D., Lappin, J. S., Gilroy, L. A., & Blake, R. (2003). Perceptual consequences of centre-surround antagonism in visual motion processing. *Nature, 424*(6946), 312–315. http://dx.doi.org/10.1038/nature01800.

Tao, X., Zhang, B., Smith, E. L., 3rd, Nishimoto, S., Ohzawa, I., & Chino, Y. M. (2012). Local sensitivity to stimulus orientation and spatial frequency within the receptive fields of neurons in visual area 2 of macaque monkeys. *Journal of Neurophysiology, 107*(4), 1094–1110. http://dx.doi.org/10.1152/jn.00640.2011.

Thoen, H. H., How, M. J., Chiou, T. H., & Marshall, J. (2014). A different form of color vision in mantis shrimp. *Science, 343*(6169), 411–413. http://dx.doi.org/10.1126/science.1245824.

Tsui, J. M., Hunter, J. N., Born, R. T., & Pack, C. C. (2010). The role of V1 surround suppression in MT motion integration. *Journal of Neurophysiology, 103*(6), 3123–3138. http://dx.doi.org/10.1152/jn.00654.2009.

Tsunoda, K., Yamane, Y., Nishizaki, M., & Tanifuji, M. (2001). Complex objects are represented in macaque inferotemporal cortex by the combination of feature columns. *Nature Neuroscience, 4*(8), 832–838. http://dx.doi.org/10.1038/90547.

van den Berg, A. V., & Noest, A. J. (1993). Motion transparency and coherence in plaids: The role of end-stopped cells. *Experimental Brain Research, 96*(3), 519–533.

van Hateren, J. H. (1992). A theory of maximizing sensory information. *Biological Cybernetics, 68*(1), 23–29.

Versavel, M., Orban, G. A., & Lagae, L. (1990). Responses of visual cortical neurons to curved stimuli and chevrons. *Vision Research, 30*(2), 235–248.

Vinje, W. E., & Gallant, J. L. (2000). Sparse coding and decorrelation in primary visual cortex during natural vision. *Science, 287*(5456), 1273–1276.

Vinje, W. E., & Gallant, J. L. (2002). Natural stimulation of the nonclassical receptive field increases information transmission efficiency in V1. *Journal of Neuroscience, 22*(7), 2904–2915. http://dx.doi.org/20026216.

von der Heydt, R., & Peterhans, E. (1989). Mechanisms of contour perception in monkey visual cortex. I. Lines of pattern discontinuity. *Journal of Neuroscience, 9*(5), 1731–1748.

Weiss, Y., Simoncelli, E. P., & Adelson, E. H. (2002). Motion illusions as optimal percepts. *Nature Neuroscience, 5*(6), 598–604. http://dx.doi.org/10.1038/nn858.

Willmore, B. D., Mazer, J. A., & Gallant, J. L. (2011). Sparse coding in striate and extrastriate visual cortex. *Journal of Neurophysiology, 105*(6), 2907–2919. http://dx.doi.org/10.1152/jn.00594.2010.

Wilson, H. R., Krupa, B., & Wilkinson, F. (2000). Dynamics of perceptual oscillations in form vision. *Nature Neuroscience, 3*(2), 170–176. http://dx.doi.org/10.1038/72115.

Wilson, H. R., Wilkinson, F., & Asaad, W. (1997). Concentric orientation summation in human form vision. *Vision Research, 37*(17), 2325–2330.

Wuerger, S., Shapley, R., & Rubin, N. (1996). "On the visually perceived direction of motion" by Hans Wallach: 60 years later. *Perception, 25*(11), 1316–1367.

Xin, D., & Bloomfield, S. A. (1999). Comparison of the responses of AII amacrine cells in the dark- and light-adapted rabbit retina. *Visual Neuroscience, 16*(4), 653–665.

Xing, D., Yeh, C. I., & Shapley, R. M. (2009). Spatial spread of the local field potential and its laminar variation in visual cortex. *Journal of Neuroscience, 29*(37), 11540–11549. http://dx.doi.org/10.1523/JNEUROSCI.2573-09.2009.

Yau, J. M., Pasupathy, A., Brincat, S. L., & Connor, C. E. (2013). Curvature processing dynamics in macaque area V4. *Cerebral Cortex, 23*(1), 198–209. http://dx.doi.org/10.1093/cercor/bhs004.

Zanos, T. P., Mineault, P. J., Monteon, J. A., & Pack, C. C. (2011). Functional connectivity during surround suppression in macaque area V4. In *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE, 2011* (pp. 3342–3345). http://dx.doi.org/10.1109/IEMBS.2011.6090906.

Zenger-Landolt, B., & Heeger, D. J. (2003). Response suppression in V1 agrees with psychophysics of surround masking. *Journal of Neuroscience, 23*(17), 6884–6893.

Zhou, H., Friedman, H. S., & von der Heydt, R. (2000). Coding of border ownership in monkey visual cortex. *Journal of Neuroscience, 20*(17), 6594–6611.

Zoccolan, D., Cox, D. D., & DiCarlo, J. J. (2005). Multiple object response normalization in monkey inferotemporal cortex. *Journal of Neuroscience, 25*(36), 8150–8164. http://dx.doi.org/10.1523/JNEUROSCI.2058-05.2005.